SUPPLEMENTARY INFORMATION FOR

Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease

Steven A. McCarroll, James E. Bradner, Hannu Turpeinen, Liisa Volin, Paul J. Martin, Shannon D. Chilewski, Joseph H. Antin , Stephanie J. Lee, Tapani Ruutu, Barry Storer, Edus H. Warren, Bo Zhang, Lue Ping Zhao, David Ginsburg, Robert J. Soiffer, Jukka Partanen, John A. Hansen, Jerome Ritz, Aarno Palotie, and David Altshuler

Contents

Supplementary Table 1

Association of donor-recipient gene-deletion mismatch (donor [–], recipient [+]) with acute GVHD. Cross-cohort analysis (including analysis of cohort A, which consisted of two cohort arms drawn from U.S and Finnish populations) utilized a Cochran-Mantel-Haenszel test; separate analyses of cohort B and cohort C utilized a chi-square test. P-values for two-sided (p2) and directional (p1) hypothesis tests are shown (see **Online Methods**); p* is corrected for multiple hypothesis testing. References document independent observations of the encoded protein as a target of a T cell clone or antibody response in individual GVHD patients.

Cohort	Gene	Refs	Test	Odds ratio [95%CI]	p2	p1	p*
A	*UGT2B17*	1–3	CMH (A1,A2)	3.2 [1.4 - 7.4]	0.01	0.005	0.03
A	*UGT2B28*		CMH (A1,A2)	1.1 [0.2 - 5.4]	0.89	0.44	
A	*GSTT1*	4	CMH (A1,A2)	0.9 [0.4 - 1.9]	0.90	0.55	
A	*GSTM1*		CMH (A1,A2)	0.9 [0.4 - 2.3]	0.89	0.56	
A	*LCE3C*		CMH (A1,A2)	0.8 [0.3 – 2.0]	0.75	0.63	
A	*OR51A2*		CMH (A1,A2)	1.1 [0.6 - 2.2]	0.93	0.47	
B	*UGT2B17*		chi-square(B)	2.3 [0.7 - 7.8]	0.17	0.09	
C	*UGT2B17*		chi-square(C)	2.4 [0.9 - 6.1]	0.07	0.04	
B + C	*UGT2B17*		CMH (B,C)	2.4 [1.1 - 5.1]	0.04	0.02	
A + B + C	*UGT2B17*		CMH(A1,A2,B,C)	2.6 [1.5 - 4.7]	0.001	0.0005	0.003

References:

1. Murata, M. *et al. J Exp Med* **197**: 1279-89 (2003).
2. Terakura, S. *et al. Transplantation* 83: 1242-8 (2005).
3. Kamei, M. *et al. Blood* **113**: 5041-8 (2009).
4. Aguilera, I. *et al. Bone Marrow Transplant*. 2009 Aug 17. (epub ahead of print).

Supplementary Table 2

Allele frequency and estimated sibling mismatch frequency for *UGT2B17* deletion in sampled populations. For all but one of the populations below, the sample is from the "extended HapMap" population sample (90 to 120 unrelated individuals per population); the Finnish population sample consists of HSC donors sampled in Helsinki (232 individuals). Allele frequency of *UGT2B17* deletion in other populations has previously been estimated by ref. 5 (with which these estimates closely agree for related populations), by ref. 6 (with which these estimates also closely agree) and by ref. 7. Under the assumption of Hardy-Weinberg equilibrium, the frequency of homozygous deletion in each population was estimated from deletion allele frequency p as p^2; sibling mismatch frequency was estimated from deletion allele frequency p as $p^2 (3+p) (1-p) / 4$.

Population sampled	Frequency of *UGT2B17* deletion allele	Frequency of homozygous deletion	Est. fraction of sibling pairs mismatched
Yoruba (Ibadan, Nigeria)	0.19	0.04	0.02
African-American (Oklahoma)	0.23	0.05	0.03
Toscani (Italy)	0.35	0.12	0.05
European ancestry (Utah)	0.35	0.12	0.05
Mexican (Los Angeles)	0.36	0.13	0.05
Luhya (Kenya)	0.38	0.14	0.06
Finnish (Helsinki)	0.48	0.23	0.08
Gujarati Indian (Houston)	0.57	0.32	0.09
Han Chinese (Beijing)	0.84	0.71	0.06
Japanese (Tokyo)	0.84	0.71	0.06
Chinese (Denver)	0.85	0.72	0.06

References:

5. Xue, Y. *et al. Am J Hum Genet* **83**: 337-46 (2008).
6. Redon R. *et al*. *Nature* **444**: 444-54 (2006).
7. Spierings E. *et al. PLoS Genet* **3**(6): e103 (2007).

Supplementary Table 3

Sequences of oligonucleotide primers and probes utilized in assays to type gene deletion polymorphisms.

PMP22 (control)	
primer1	CCCTTCTCAGCGGTGTCATC
primer2	ACAGACCGTCTGGGCGC
probe	VIC - TTCGCGTTTCCGCAAGAT – MGBNFQ
UGT2B17	
primer1	AAGACGTTTTGTCGCAGGAA
primer2	GCCTGAAGTGGAATGACCAA
probe	FAM - CCCTCCATGCTGGAATAAAGGAGGA - BHQ1
UGT2B28	
primer1	CAGGTGGTCAGCTTCAGAGA
primer2	ATGTTTTGAAGGTGGGAAGC
probe	FAM - TGCAGGCTCAGCTCTGCAGATG - BHQ1
LCE3C	
primer1	ACAAACAGAGGAGCGAGGAA
primer2	TCCATACCCATCCTGGTGAT
probe	FAM - AGCCCCTCATGGAGGGGAGG - BHQ1
GSTM1	
primer1	CTGTGTCCACCTGCATTCG
primer2	GAGACCGGGCACTCACTGT
probe	FAM – TCAGTCCTGCCATGAGCAGGC - BHQ1
GSTT1	
primer1	GGGATGGAAAGTCACGTCCT
primer2	AGAGACTGGGACAGCGTCAA
probe	FAM – CAGAATCTCAGCAGCTGGGCCA - BHQ1
OR51A2	
primer1	TGCCAATTGCCTACTGTTTG
primer2	AGCAACAGTGGAAGGAGAGAA
probe	FAM - TGACAACATAACCAAGTGGGGCTTATTTTC- BHQ1

UGT2B17	Replication assay
primer1	CCTTCACATGCACATTGGTC
primer2	CATGCAGATTTTCCCCTGTT
probe	HEX- AGGCTTCCCTGGGAGCCCAG – BHQ1

Supplementary Note

Genotyping of gene deletion polymorphisms

1. Genotyping of gene deletion polymorphisms by quantitative PCR

Deletions were typed using a two-color TaqMan assay. To type each deletion, the locus of interest and a control, two-copy locus were amplified simultaneously in each reaction well. Simultaneous amplification of the two genomic segments in the same reaction well was detected in real time using FAM and VIC fluorophores respectively. Comparison of the amplification curves for the two fluorophores offers an internally controlled measurement of the relative copy number of the two genomic segments.

1A. Primers and probes

Oligonucleotide primers and FAM-labeled probes were obtained from Integrated DNA Technologies. VIC-labeled control probes were obtained from Applied Biosystems. Primer and probe sequences are shown in Supplementary Table 3.

1B. Experimental setup and cycling conditions

Samples were assayed in 20ul reactions containing 1X TaqMan master mix (Applied Biosystems), 125 nM of each primer and probe, and 20 ng genomic DNA. Reactions were prepared in 384-well plates. Reactions were thermocyled in an Applied Biosystems 7900 HT instrument using the following cycling protocol:

1. 2 minutes at 50C; then
2. 10 minutes at 95C; then
3. 40 cycles of (15 seconds at 95C, 90 seconds at 56C)

Real-time data was collected from the FAM and VIC channels during each amplification cycle. A threshold amplification cycle (*Ct*) was calculated for each fluorophore in each reaction well using the Applied Biosystems SDS 3.0 software with default parameters.

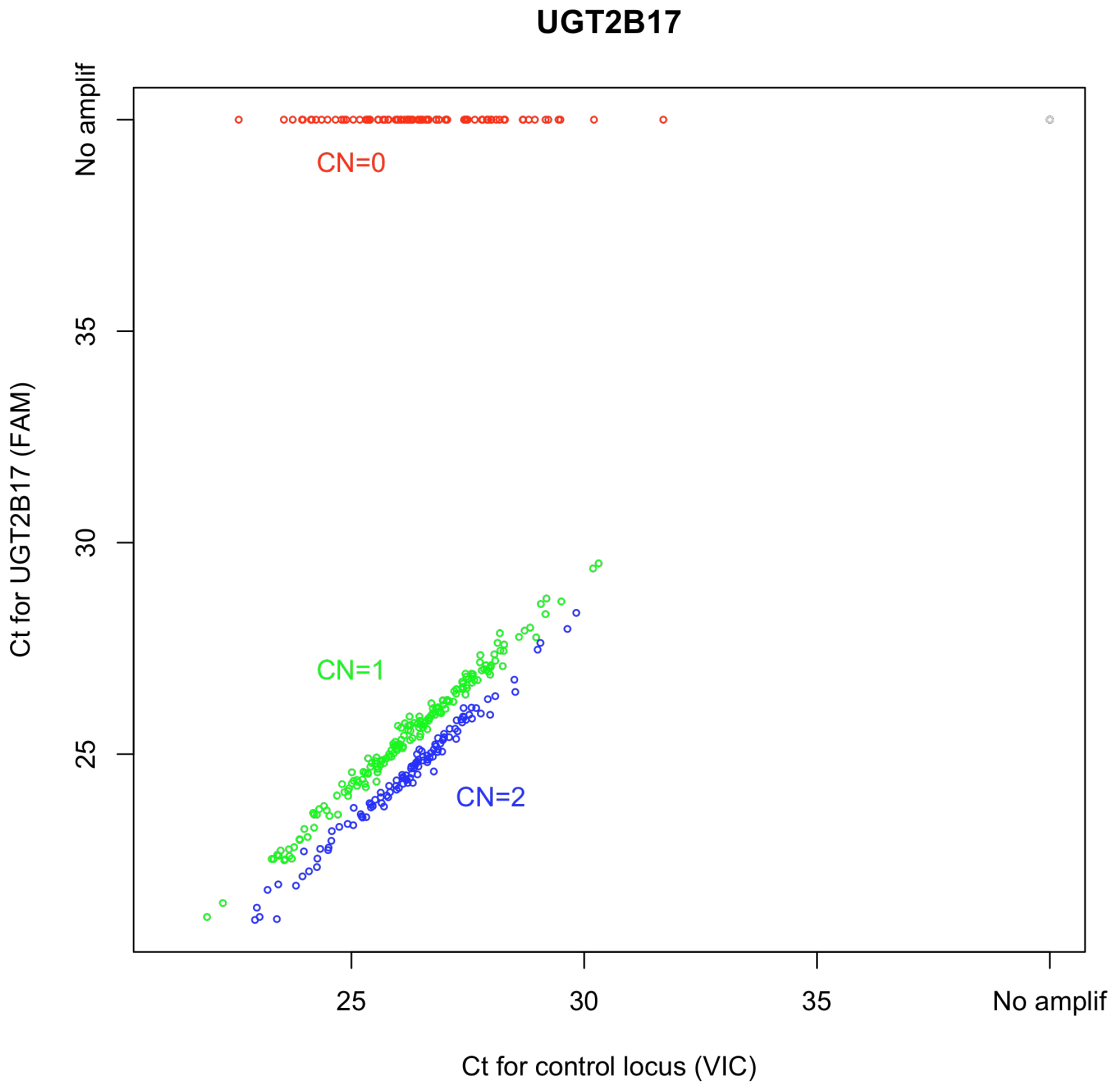
1C. Analysis of data and derivation of genotypes

The two-color threshold-cycle data for each sample is readily visualized on a scatter plot showing the data for a large number of samples. On the following pages, each scatter plot shows (for one assay in one 384-well plate) the threshold-cycle (*Ct*) measurements derived for the FAM and VIC channels (corresponding, respectively, to the locus of interest and the control locus). Each point corresponds to one DNA sample. Samples are colored according to the genotype call (red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control).

Samples were determined to have a homozygous deletion (CN=0, red) if, in the same reaction well:

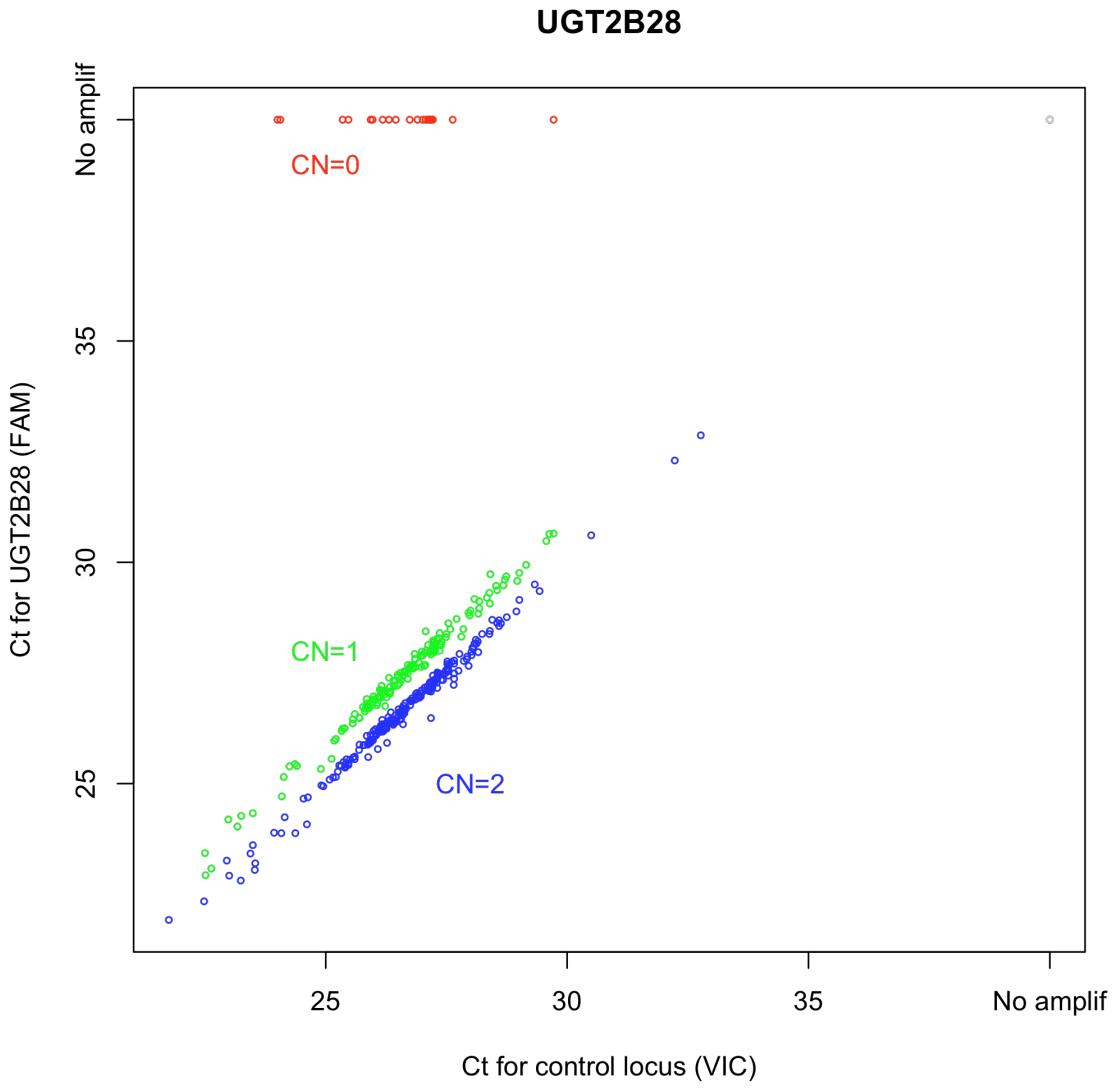
1. The control amplicon showed robust amplification, and
2. The assay amplicon (for the gene-deletion locus) showed no amplification.

Distinction between the other two copy-number classes (CN=1 and CN=2), while not germane to the hypothesis in this study, was important for scrutinizing the overall quality of the deletion genotypes: it allowed us to test the conformity of genotypes to Hardy-Weinberg equilibrium and to the expected level of allele sharing between siblings (see section 1D below). To assign samples to the CN1 and CN2 classes, the quantity ($Ct^{VIC} - Ct^{FAM}$) was calculated and clustered, in each case showing a multimodal distribution that allowed resolution of the copy-number classes.



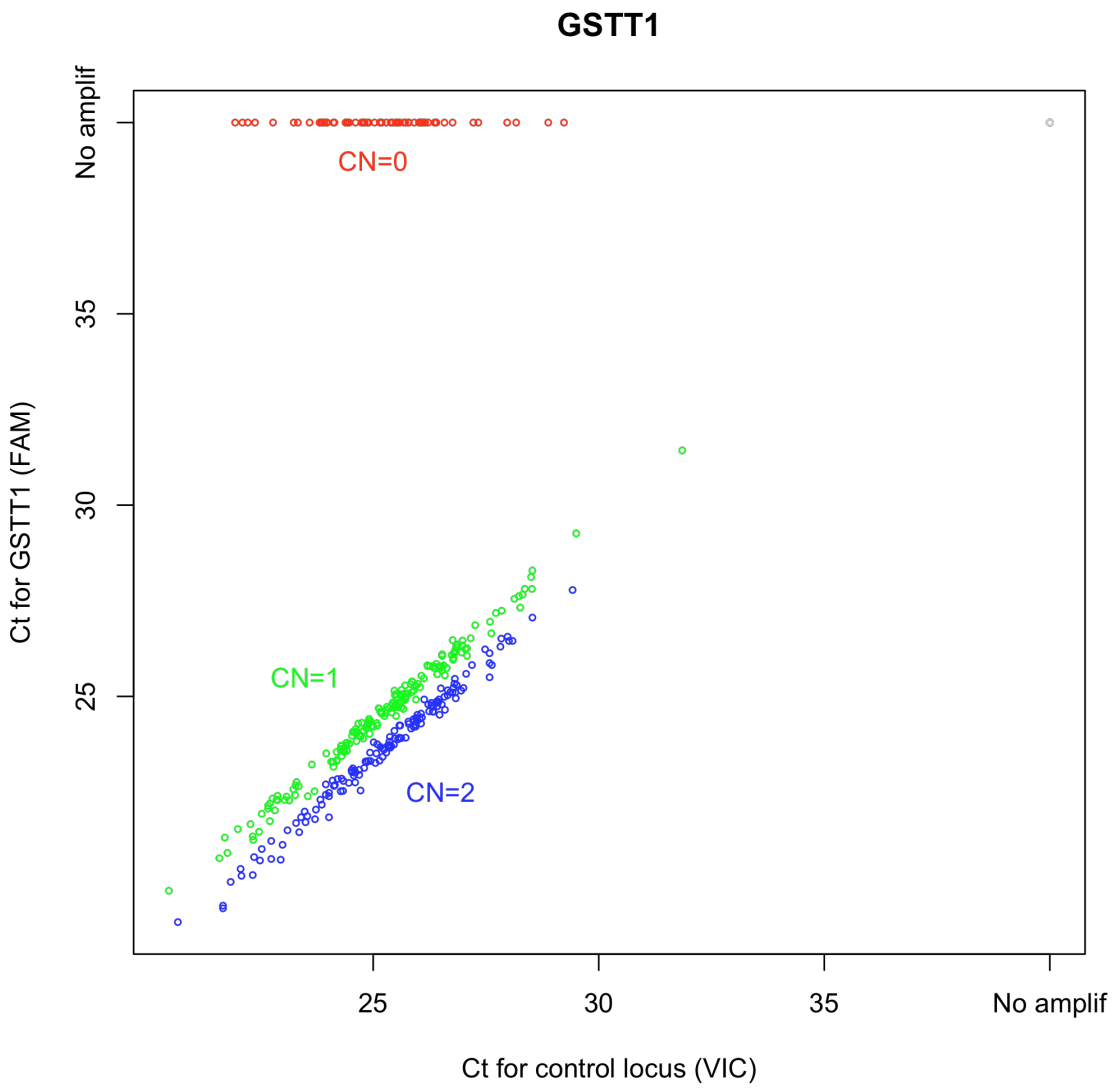
Deletion genotype assay for *UGT2B17*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe. Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe. Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.



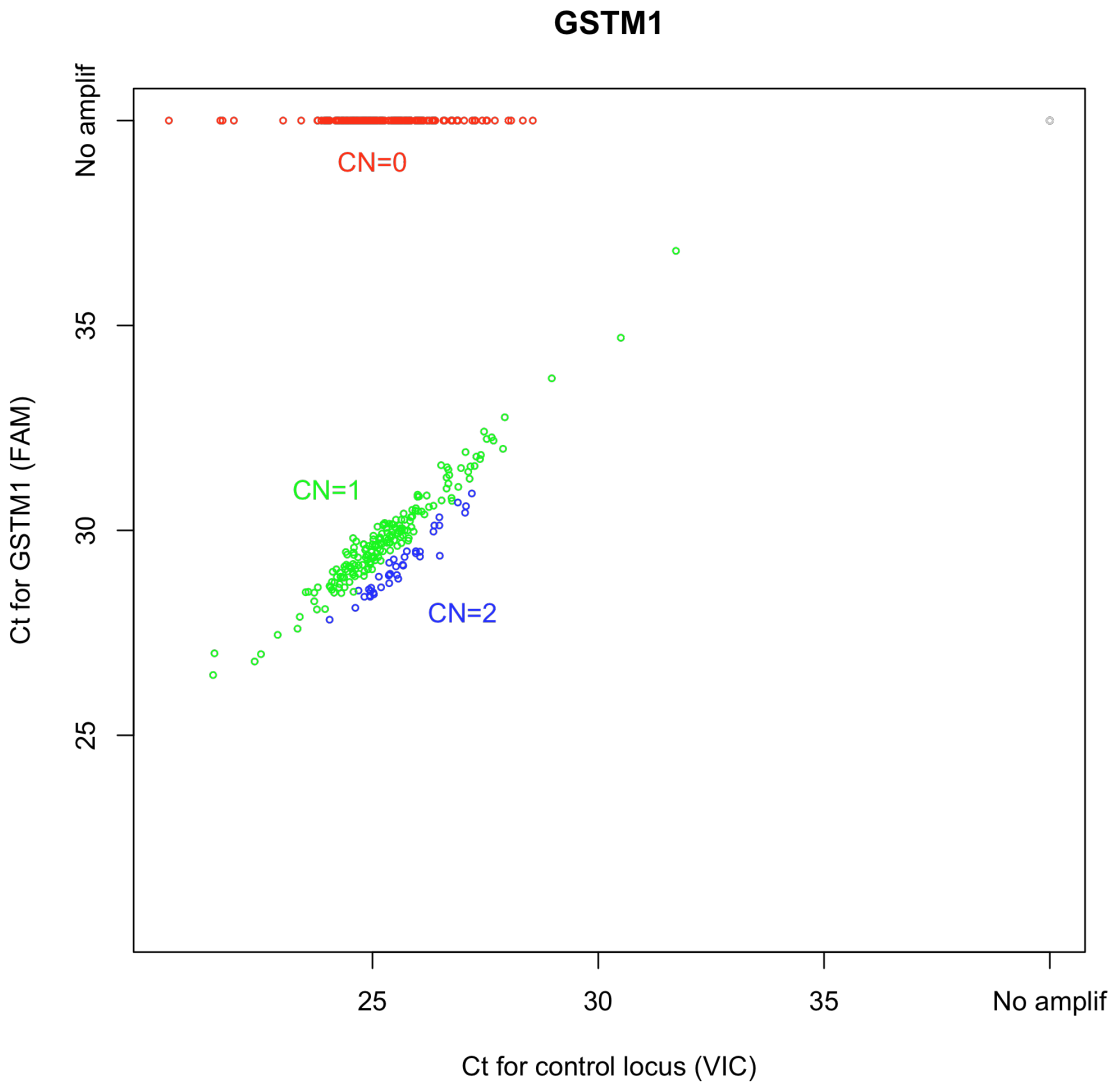
Deletion genotype assay for *UGT2B28*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe
Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe
Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.



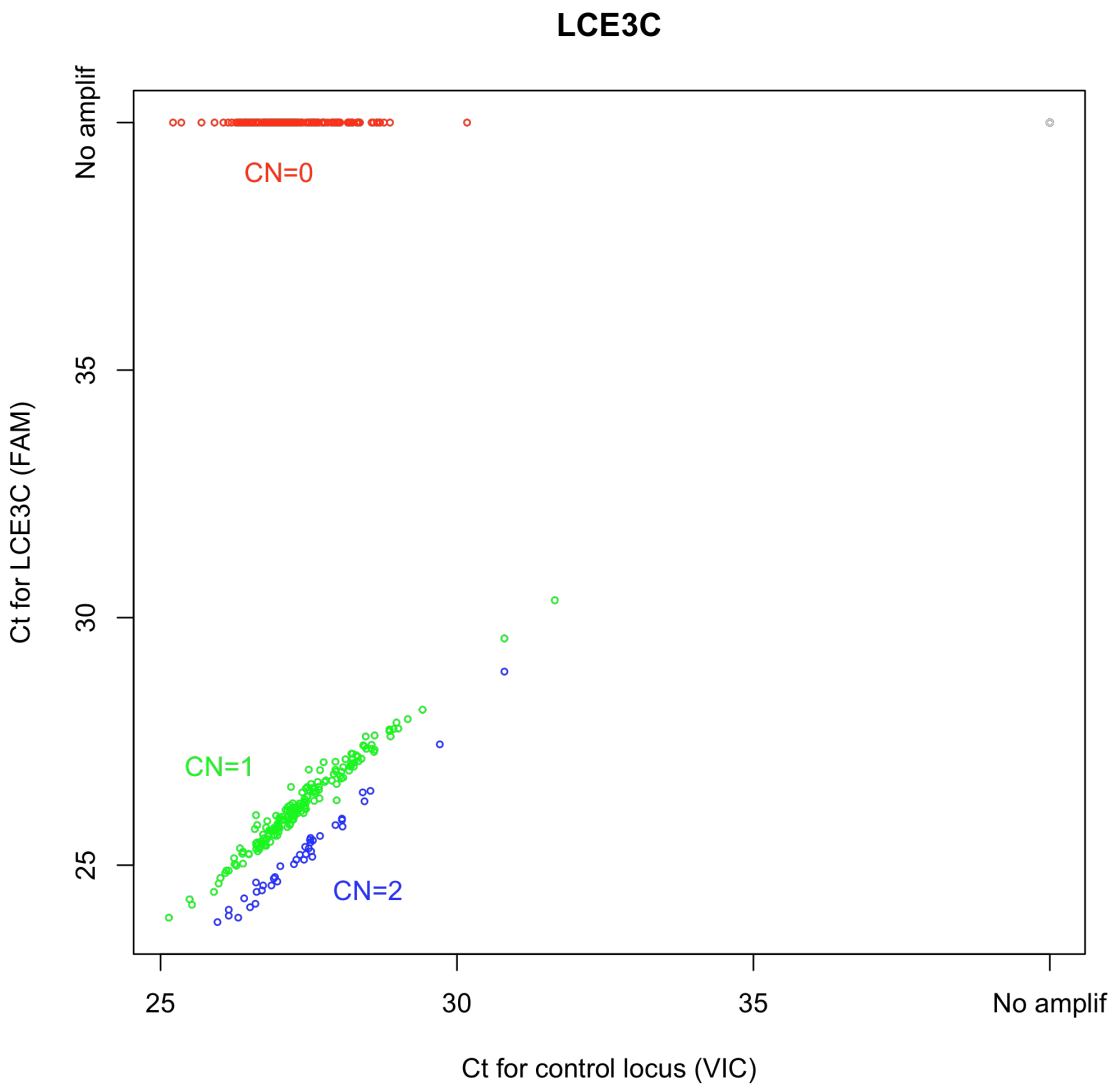
Deletion genotype assay for *GSTT1*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe
Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe
Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.



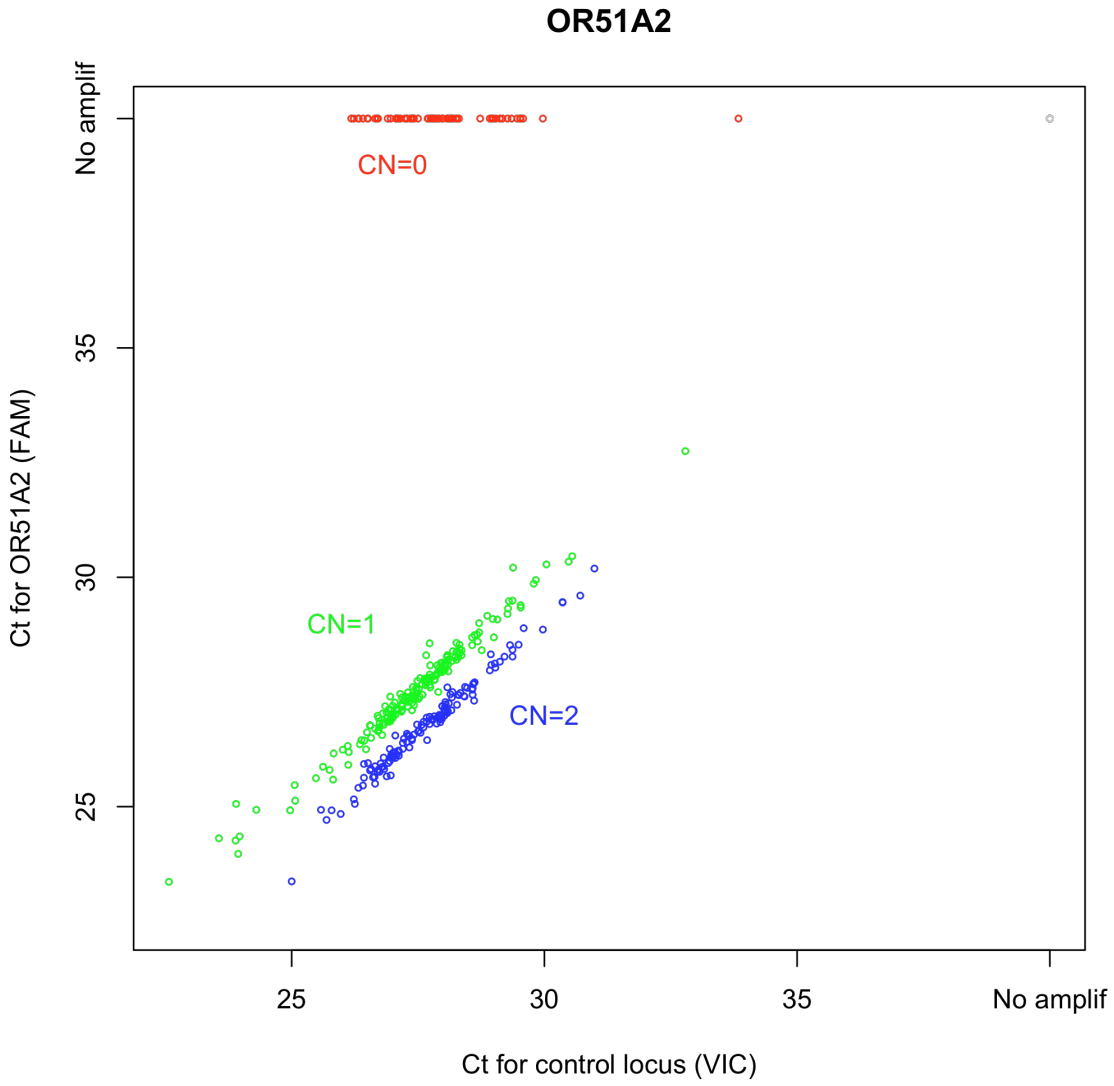
Deletion genotype assay for *GSTM1*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe
Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe
Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.



Deletion genotype assay for *LCE3C*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe
Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe
Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.



Deletion genotype assay for *OR51A2*

Horizontal axis: Threshold cycle (Ct) for amplification of control locus, detected using VIC-labeled probe
Vertical axis: Threshold cycle (Ct) for amplification of test locus, detected using FAM-labeled probe
Each point corresponds to one patient sample. Data for 384 samples (assayed on the same 384-well plate) are shown. Each point is colored by the called deletion genotype: red = homozygous deletion / 0 copies; green = heterozygous deletion / 1 copy; blue = no deletion / 2 copies; gray = no-DNA control.

1D. Additional quality-control checks

For each of the six gene-deletion assays and each of the four clinical populations, we assessed the conformance of the resulting genotypes to Hardy-Weinberg equilibrium, using a p-value threshold of 0.01; 0/24 genotype sets violated Hardy-Weinberg equilibrium.

We also evaluated the conformance of each set of genotypes to the expected level of allele sharing between siblings (50% on average) by assessing the correlation of transplant-recipient genotypes to the genotypes of their sibling donors. The correlation coefficient (r) between donor and recipient genotypes was not significantly (p=0.01) different from 0.50 in any population for any assay in any population sample (0/24).

Finally, we assessed the concordance (across the 270 HapMap samples) of the quantitative PCR assays with data from an independent approach, the Affymetrix SNP6.0 array, on which we genotyped these same six deletion polymorphisms (as described in McCarroll et al., *Nature Genetics* 2008), with the following concordance observed:

	Concordance (CN=0 vs CN>0)	Concordance (CN=1 vs CN=2)
UGT2B17	100%	98.8%
UGT2B28	100%	99.4%
GSTT1	100%	98.5%
GSTM1	100%	96.2%
LCE3C	100%	99.5%
OR51A2	100%	99.5%

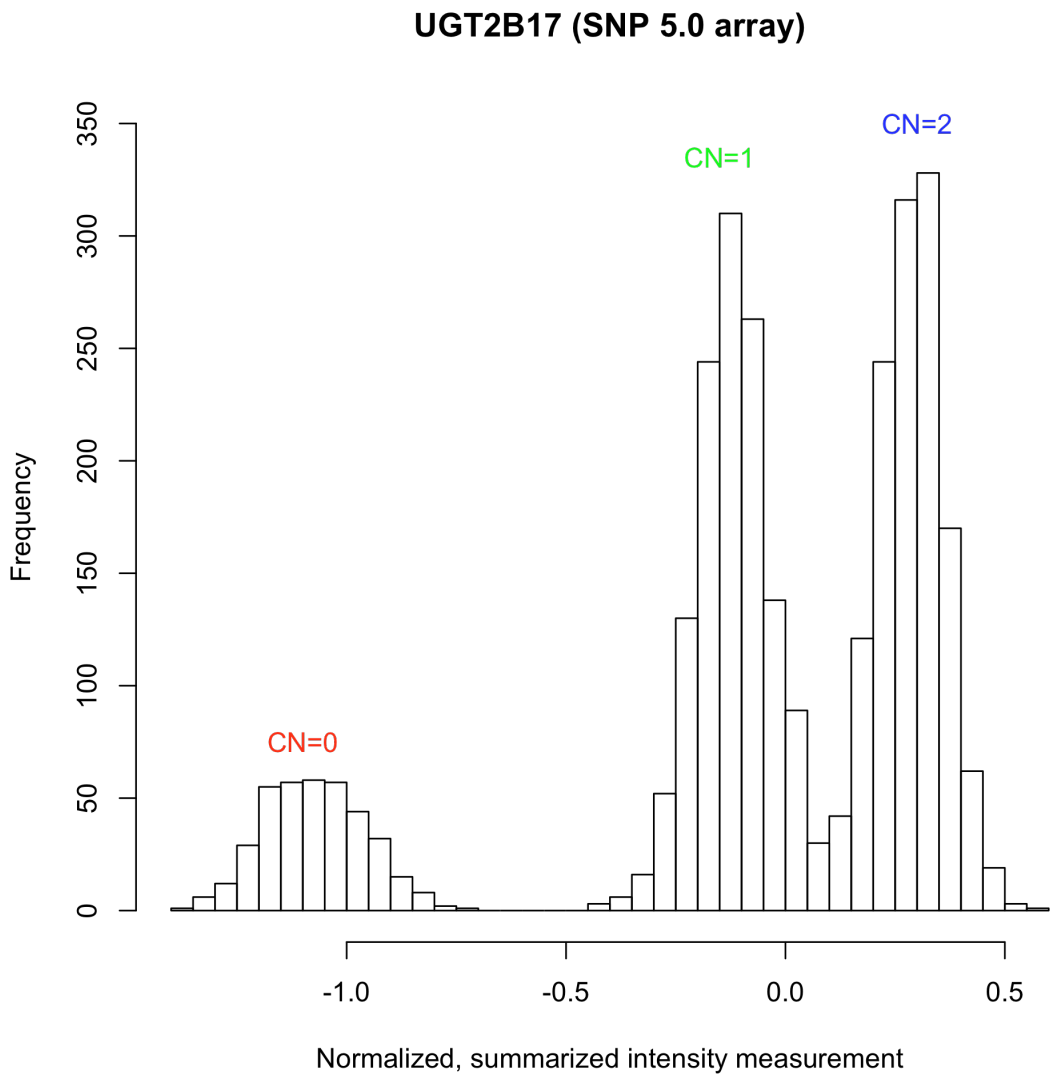
1E. Confirmatory quantitative PCR assay for UGT2B17

To confirm genotypes for *UGT2B17,* we developed a second quantitative PCR assay, taking the following steps to make it independent of the first assay: (1) we utilized a different amplification site within the *UGT2B17* deletion, 40 kb away (on the hg36 reference sequence) from the site interrogated in the initial assay; (2) we utilized a different control locus, on a different chromosome from the control locus in the first assay; (3) we reversed the fluorophores on the two probes, such that the control locus probe was labeled with FAM and the test locus probe was labeled with HEX (**Supplementary Table 3**).

Results for the additional assay were 100.0% concordant with the first quantitative PCR assay in distinguishing CN=0 from CN>0 individuals and 99.2% concordant in distinguishing CN=1 from CN=2.

2. Genotyping of UGT2B17 deletion polymorphism using SNP 5.0 array

One Affymetrix SNP 5.0 array was run on each subject. Across the 120-kb genomic region spanned by the *UGT2B17* deletion, the Affymetrix SNP 5.0 array contains 20 "copy number" probes – probes that interrogate non-polymorphic sequences and that are optimized for copy number determination (McCarroll et al., *Nat Genet* **40**, 1166-74 (2008)). We first performed quantile normalization to normalize raw intensity values across arrays. The normalized intensity measurement for each probe was then divided by the population median measurement (for that probe) to obtain a log-ratio. Measurements were mean-summarized across the 20 probes that interrogate the genomic region affected by the deletion. The resulting, summarized measurements (one per sample) showed a trimodal distribution corresponding to the three copy-number genotype classes CN0 (homozygous deletion), CN1, and CN2.



To critically evaluate the accuracy of deletion genotypes obtained in this manner, we repeated this analysis on SNP5.0 data for 270 HapMap samples (McCarroll et al., *Nat Genet* **40**, 1166-74 (2008)) and also performed the quantitative PCR assay (described above) on these same 270

samples. Data from the two assays showed 100.0% concordance on CN0 genotypes and 97.8% concordance in distinguishing CN1 from CN2 genotypes.

UGT2B17 deletion genotypes from the SNP5.0 array also passed the additional quality-control tests imposed on deletion genotypes derived from quantitative PCR (section 1D, above).